

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.**

Application Serial Number: 10/594,731  
Source: 1 FWP  
Date Processed by STIC: 10/06/06

# ***ENTERED***



IFWP

## RAW SEQUENCE LISTING

DATE: 10/06/2006

PATENT APPLICATION: US/10/594,731

TIME: 11:13:17

Input Set : A:\1034232.APP

Output Set: N:\CRF4\10062006\J594731.raw

```

3 <110> APPLICANT: MATSUMOTO, KAZUYA
4     KAZUNO, YASUSHI
5     HIGASHIMURA, NORIKAZU
6     OHSHIMA, TOSHIHISA
7     SAKURABA, HARUHIKO
9 <120> TITLE OF INVENTION: PROCESS FOR PRODUCING CHIRAL HYDROXYALDEHYDE COMPOUNDS
11 <130> FILE REFERENCE: 1034232-000045
C--> 13 <140> CURRENT APPLICATION NUMBER: US/10/594,731
C--> 14 <141> CURRENT FILING DATE: 2006-09-29
16 <150> PRIOR APPLICATION NUMBER: PCT/JP2005/005719
17 <151> PRIOR FILING DATE: 2005-03-28
19 <150> PRIOR APPLICATION NUMBER: JP 2004-095263
20 <151> PRIOR FILING DATE: 2004-03-29
22 <160> NUMBER OF SEQ ID NOS: 10
24 <170> SOFTWARE: PatentIn Ver. 3.3
26 <210> SEQ ID NO: 1
27 <211> LENGTH: 747
28 <212> TYPE: DNA
29 <213> ORGANISM: Thermotoga maritima
31 <220> FEATURE:
32 <221> NAME/KEY: CDS
33 <222> LOCATION: (1)..(744)
35 <400> SEQUENCE: 1
36 atg ata gag tac agg att gag gag gca gta gcg aag tac aga gag ttc      48
37 Met Ile Glu Tyr Arg Ile Glu Glu Ala Val Ala Lys Tyr Arg Glu Phe
38   1           5           10           15
40 tac gaa ttc aag ccc gtc aga gaa agc gca ggt att gaa gat gtg aaa      96
41 Tyr Glu Phe Lys Pro Val Arg Glu Ser Ala Gly Ile Glu Asp Val Lys
42   20           25           30
44 agt gct ata gag cac acg aat ctg aaa ccg ttt gcc aca cca gac gat      144
45 Ser Ala Ile Glu His Thr Asn Leu Lys Pro Phe Ala Thr Pro Asp Asp
46   35           40           45
48 ata aaa aaa ctc tgt ctt gaa gca agg gaa aat cgt ttc cat gga gtc      192
49 Ile Lys Lys Leu Cys Leu Glu Ala Arg Glu Asn Arg Phe His Gly Val
50   50           55           60
52 tgt gtg aat ccg tgt tat gtg aaa ctg gct cgt gaa gaa ctc gaa gga      240
53 Cys Val Asn Pro Cys Tyr Val Lys Leu Ala Arg Glu Glu Leu Glu Gly
54   65           70           75           80
56 acc gat gtg aaa gtc gtc acc gtt gtt ggt ttt cca ctg gga gcg aac      288
57 Thr Asp Val Lys Val Val Thr Val Val Gly Phe Pro Leu Gly Ala Asn
58   85           90           95
60 gaa act cgg acg aaa gcc cat gag gcg att ttc gct gtt gag agt gga      336
61 Glu Thr Arg Thr Lys Ala His Glu Ala Ile Phe Ala Val Glu Ser Gly

```

## RAW SEQUENCE LISTING

DATE: 10/06/2006

PATENT APPLICATION: US/10/594,731

TIME: 11:13:17

Input Set : A:\1034232.APP

Output Set: N:\CRF4\10062006\J594731.raw

```

62          100          105          110
64 gcc gat gag atc gat atg gtc atc aac gtt ggc atg ctc aag gca aag 384
65 Ala Asp Glu Ile Asp Met Val Ile Asn Val Gly Met Leu Lys Ala Lys
66          115          120          125
68 gag tgg gag tac gtt tac gag gat ata aga agt gtt gtc gaa tcg gtg 432
69 Glu Trp Glu Tyr Val Tyr Glu Asp Ile Arg Ser Val Val Glu Ser Val
70          130          135          140
72 aaa gga aaa gtt gtg aag gtg atc atc gaa acg tgc tat ctg gat acg 480
73 Lys Gly Lys Val Val Lys Val Ile Ile Glu Thr Cys Tyr Leu Asp Thr
74 145          150          155          160
76 gaa gag aag ata gcg gcg tgt gtc att tcc aaa ctt gct gga gct cat 528
77 Glu Glu Lys Ile Ala Ala Cys Val Ile Ser Lys Leu Ala Gly Ala His
78          165          170          175
80 ttc gtg aag act tcc acg gga ttt gga aca gga ggg gcg acc gca gaa 576
81 Phe Val Lys Thr Ser Thr Gly Phe Gly Thr Gly Gly Ala Thr Ala Glu
82          180          185          190
84 gac gtt cat ctc atg aaa tgg atc gtg gga gat gag atg ggt gta aaa 624
85 Asp Val His Leu Met Lys Trp Ile Val Gly Asp Glu Met Gly Val Lys
86          195          200          205
88 gct tcc gga ggg atc aga acc ttc gag gac gct gtt aaa atg atc atg 672
89 Ala Ser Gly Gly Ile Arg Thr Phe Glu Asp Ala Val Lys Met Ile Met
90          210          215          220
92 tac ggt gct gat aga ata gga acg agt tcg gga gtt aag atc gtt cag 720
93 Tyr Gly Ala Asp Arg Ile Gly Thr Ser Ser Gly Val Lys Ile Val Gln
94 225          230          235          240
96 ggg gga gaa gag aga tat gga ggt tga 747
97 Gly Gly Glu Glu Arg Tyr Gly Gly
98          245
101 <210> SEQ ID NO: 2
102 <211> LENGTH: 248
103 <212> TYPE: PRT
104 <213> ORGANISM: Thermotoga maritima
106 <400> SEQUENCE: 2
107 Met Ile Glu Tyr Arg Ile Glu Glu Ala Val Ala Lys Tyr Arg Glu Phe
108 1 5 10 15
110 Tyr Glu Phe Lys Pro Val Arg Glu Ser Ala Gly Ile Glu Asp Val Lys
111 20 25 30
113 Ser Ala Ile Glu His Thr Asn Leu Lys Pro Phe Ala Thr Pro Asp Asp
114 35 40 45
116 Ile Lys Lys Leu Cys Leu Glu Ala Arg Glu Asn Arg Phe His Gly Val
117 50 55 60
119 Cys Val Asn Pro Cys Tyr Val Lys Leu Ala Arg Glu Glu Leu Glu Gly
120 65 70 75 80
122 Thr Asp Val Lys Val Val Thr Val Val Gly Phe Pro Leu Gly Ala Asn
123 85 90 95
125 Glu Thr Arg Thr Lys Ala His Glu Ala Ile Phe Ala Val Glu Ser Gly
126 100 105 110
128 Ala Asp Glu Ile Asp Met Val Ile Asn Val Gly Met Leu Lys Ala Lys
129 115 120 125

```

## RAW SEQUENCE LISTING

DATE: 10/06/2006

PATENT APPLICATION: US/10/594,731

TIME: 11:13:17

Input Set : A:\1034232.APP

Output Set: N:\CRF4\10062006\J594731.raw

```

131 Glu Trp Glu Tyr Val Tyr Glu Asp Ile Arg Ser Val Val Glu Ser Val
132      130      135      140
134 Lys Gly Lys Val Val Lys Val Ile Ile Glu Thr Cys Tyr Leu Asp Thr
135 145      150      155      160
137 Glu Glu Lys Ile Ala Ala Cys Val Ile Ser Lys Leu Ala Gly Ala His
138      165      170      175
140 Phe Val Lys Thr Ser Thr Gly Phe Gly Thr Gly Gly Ala Thr Ala Glu
141      180      185      190
143 Asp Val His Leu Met Lys Trp Ile Val Gly Asp Glu Met Gly Val Lys
144      195      200      205
146 Ala Ser Gly Gly Ile Arg Thr Phe Glu Asp Ala Val Lys Met Ile Met
147      210      215      220
149 Tyr Gly Ala Asp Arg Ile Gly Thr Ser Ser Gly Val Lys Ile Val Gln
150 225      230      235      240
152 Gly Gly Glu Glu Arg Tyr Gly Gly
153      245
157 <210> SEQ ID NO: 3
158 <211> LENGTH: 681
159 <212> TYPE: DNA
160 <213> ORGANISM: Pyrobaculum aerophilum
162 <220> FEATURE:
163 <221> NAME/KEY: CDS
164 <222> LOCATION: (1)..(678)
166 <400> SEQUENCE: 3
167 atg ata cat tta gta gac tac gcg ctt ctc aag ccg tat ctc aca gta 48
168 Met Ile His Leu Val Asp Tyr Ala Leu Leu Lys Pro Tyr Leu Thr Val
169 1 5 10 15
171 gat gaa gca gtc gcc ggg gct cgc aag gcg gag gag ctg ggc gtc gcg 96
172 Asp Glu Ala Val Ala Gly Ala Arg Lys Ala Glu Glu Leu Gly Val Ala
173 20 25 30
175 gcg tat tgc gta aat ccc ata tac gcc cct gtt gtt cgg cct ttg ttg 144
176 Ala Tyr Cys Val Asn Pro Ile Tyr Ala Pro Val Val Arg Pro Leu Leu
177 35 40 45
179 cgg aaa gta aag ctc tgc gta gtg gcg gac ttc ccc ttt ggg gcc ttg 192
180 Arg Lys Val Lys Leu Cys Val Val Ala Asp Phe Pro Phe Gly Ala Leu
181 50 55 60
183 cca acg gcc agc aga att gcc ttg gtt tct agg ctt gct gaa gtg gca 240
184 Pro Thr Ala Ser Arg Ile Ala Leu Val Ser Arg Leu Ala Glu Val Ala
185 65 70 75 80
187 gat gag ata gac gtg gtg gcg cct ata ggc ctc gtg aaa tcg cgg agg 288
188 Asp Glu Ile Asp Val Val Ala Pro Ile Gly Leu Val Lys Ser Arg Arg
189 85 90 95
191 tgg gcc gag gtg aga agg gac tta ata agc gtt gtg ggt gcc gca ggc 336
192 Trp Ala Glu Val Arg Arg Asp Leu Ile Ser Val Val Gly Ala Ala Gly
193 100 105 110
195 ggg aga gtg gta aag gta atc aca gag gag cct tat cta agg gat gag 384
196 Gly Arg Val Val Lys Val Ile Thr Glu Glu Pro Tyr Leu Arg Asp Glu
197 115 120 125
199 gag agg tat acg ctt tac gac att att gca gag gct ggg gcc cac ttt 432

```

## RAW SEQUENCE LISTING

DATE: 10/06/2006

PATENT APPLICATION: US/10/594,731

TIME: 11:13:17

Input Set : A:\1034232.APP

Output Set: N:\CRF4\10062006\J594731.raw

```

200 Glu Arg Tyr Thr Leu Tyr Asp Ile Ile Ala Glu Ala Gly Ala His Phe
201      130                      135                      140
203 ata aaa agc tcc act gga ttc gcc gaa gag gcc tac gcc gcc aga cag 480
204 Ile Lys Ser Ser Thr Gly Phe Ala Glu Glu Ala Tyr Ala Ala Arg Gln
205 145                      150                      155                      160
207 gga aat cct gta cac tca acg ccg gag agg gcg gcg gca att gcc cgc 528
208 Gly Asn Pro Val His Ser Thr Pro Glu Arg Ala Ala Ala Ile Ala Arg
209                      165                      170                      175
211 tac ata aaa gag aag ggg tat aga ctg ggg gtg aaa atg gcg ggg ggg 576
212 Tyr Ile Lys Glu Lys Gly Tyr Arg Leu Gly Val Lys Met Ala Gly Gly
213                      180                      185                      190
215 att agg aca agg gag cag gca aag gcc att gtt gac gcc att gga tgg 624
216 Ile Arg Thr Arg Glu Gln Ala Lys Ala Ile Val Asp Ala Ile Gly Trp
217                      195                      200                      205
219 ggc gag gac cca gcc cgc gtc agg ctg ggg acg tcc acc cca gag gct 672
220 Gly Glu Asp Pro Ala Arg Val Arg Leu Gly Thr Ser Thr Pro Glu Ala
221      210                      215                      220
223 ctt cta tag 681
224 Leu Leu
225 225
228 <210> SEQ ID NO: 4
229 <211> LENGTH: 226
230 <212> TYPE: PRT
231 <213> ORGANISM: Pyrobaculum aerophilum
233 <400> SEQUENCE: 4
234 Met Ile His Leu Val Asp Tyr Ala Leu Leu Lys Pro Tyr Leu Thr Val
235      1                      5                      10                      15
237 Asp Glu Ala Val Ala Gly Ala Arg Lys Ala Glu Glu Leu Gly Val Ala
238                      20                      25                      30
240 Ala Tyr Cys Val Asn Pro Ile Tyr Ala Pro Val Val Arg Pro Leu Leu
241                      35                      40                      45
243 Arg Lys Val Lys Leu Cys Val Val Ala Asp Phe Pro Phe Gly Ala Leu
244      50                      55                      60
246 Pro Thr Ala Ser Arg Ile Ala Leu Val Ser Arg Leu Ala Glu Val Ala
247 65                      70                      75                      80
249 Asp Glu Ile Asp Val Val Ala Pro Ile Gly Leu Val Lys Ser Arg Arg
250                      85                      90                      95
252 Trp Ala Glu Val Arg Arg Asp Leu Ile Ser Val Val Gly Ala Ala Gly
253                      100                      105                      110
255 Gly Arg Val Val Lys Val Ile Thr Glu Glu Pro Tyr Leu Arg Asp Glu
256      115                      120                      125
258 Glu Arg Tyr Thr Leu Tyr Asp Ile Ile Ala Glu Ala Gly Ala His Phe
259      130                      135                      140
261 Ile Lys Ser Ser Thr Gly Phe Ala Glu Glu Ala Tyr Ala Ala Arg Gln
262 145                      150                      155                      160
264 Gly Asn Pro Val His Ser Thr Pro Glu Arg Ala Ala Ala Ile Ala Arg
265                      165                      170                      175
267 Tyr Ile Lys Glu Lys Gly Tyr Arg Leu Gly Val Lys Met Ala Gly Gly
268                      180                      185                      190

```

## RAW SEQUENCE LISTING

DATE: 10/06/2006

PATENT APPLICATION: US/10/594,731

TIME: 11:13:17

Input Set : A:\1034232.APP

Output Set: N:\CRF4\10062006\J594731.raw

```

270 Ile Arg Thr Arg Glu Gln Ala Lys Ala Ile Val Asp Ala Ile Gly Trp
271      195                      200                      205
273 Gly Glu Asp Pro Ala Arg Val Arg Leu Gly Thr Ser Thr Pro Glu Ala
274      210                      215                      220
276 Leu Leu
277 225
281 <210> SEQ ID NO: 5
282 <211> LENGTH: 30
283 <212> TYPE: DNA
284 <213> ORGANISM: Artificial Sequence
286 <220> FEATURE:
287 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
288     Primer
290 <400> SEQUENCE: 5
291 tatatcatat gatagagtac aggattgagg                      30
294 <210> SEQ ID NO: 6
295 <211> LENGTH: 30
296 <212> TYPE: DNA
297 <213> ORGANISM: Artificial Sequence
299 <220> FEATURE:
300 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
301     Primer
303 <400> SEQUENCE: 6
304 taatggatcc tcaacctcca tatctctctt                      30
307 <210> SEQ ID NO: 7
308 <211> LENGTH: 30
309 <212> TYPE: DNA
310 <213> ORGANISM: Artificial Sequence
312 <220> FEATURE:
313 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
314     Primer
316 <400> SEQUENCE: 7
317 tatatcatat gatacattta gtagactacg                      30
320 <210> SEQ ID NO: 8
321 <211> LENGTH: 30
322 <212> TYPE: DNA
323 <213> ORGANISM: Artificial Sequence
325 <220> FEATURE:
326 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
327     Primer
329 <400> SEQUENCE: 8
330 taatggatcc ctatagaaga gcctctgagg                      30
333 <210> SEQ ID NO: 9
334 <211> LENGTH: 30
335 <212> TYPE: DNA
336 <213> ORGANISM: Artificial Sequence
338 <220> FEATURE:
339 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
340     Primer

```

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/10/594,731

DATE: 10/06/2006

TIME: 11:13:18

Input Set : A:\1034232.APP

Output Set: N:\CRF4\10062006\J594731.raw

L:13 M:270 C: Current Application Number differs, Replaced Application Number

L:14 M:271 C: Current Filing Date differs, Replaced Current Filing Date